Solving Support Vector Machines in Reproducing Kernel Banach Spaces with Positive Definite Functions

Gregory E. Fasshauer, Fred J. Hickernell and Qi Ye*
Communicated by Yuesheng Xu

Abstract

In this paper we extend support vector machines from reproducing kernel Hilbert spaces into reproducing kernel Banach spaces whose reproducing kernels can be defined on nonsymmetric domains. Using the orthogonality of semi-inner products, we can obtain the empirical representations of support vector machine solutions. In addition, we can set up the reproduction property onto a generalized native space by Fourier transform techniques in order that it becomes a reproducing kernel Banach space and its reproducing kernel is given by the related positive definite function. The reproducing kernel Banach spaces of some reproducing kernels can be even imbedded into Sobolev spaces. We show some special examples of reproducing kernel Banach spaces induced by Matérn functions (Sobolev splines) such that their support vector machine solutions are well computable same as the classical cases but their explicit formulas are totally different from the solutions on reproducing kernel Hilbert spaces. It is possible to produce a new numerical tool for support vector machines.

Keywords: support vector machine, reproducing kernel Banach space, reproducing kernel, positive definite function, Fourier transform, Sobolev space, Matérn function, Sobolev spline.

1 Introduction

The theory and practice of kernel-based methods is a fast growing research area. It has been used for both scattered data approximation and machine learning. Their applications come from such different fields as physics, biology, geology, meteorology and finance. The books [4, 7, 21] show how to use (conditionally) positive definite kernels to construct interpolants for given data sites sampled from some unknown functions in the native spaces corresponded to the kernel functions. In the books [2, 3, 19, 20], we can choose the optimal support vector machine solution in a reproducing kernel Hilbert space (RKHS) of a given reproducing kernel and this solution is introduced by the related reproducing kernel and the practical data values. Actually, the native spaces and the RKHSs are in the same senses and the numerical people and the statical learning people just use different languages and techniques to introduce

*Corresponding author

the spaces. Moreover, the recent papers [9, 10, 22] even develop a clear and detailed framework of generalized Sobolev spaces and RKHSs to make a connection with their associated Green functions and reproducing kernels.

In the current research work, the papers [5, 6] extend classical native spaces to generalized native spaces and the papers [17, 18, 23] generalize RKHSs to Banach spaces. However, the reproductions of generalized native spaces are not discussed in [5, 6] and [17, 18, 23] do not mention how to use reproducing kernels to set up the explicit forms of the related reproducing kernel Banach spaces (RKBSs) defined on continuous domains, whose structure is difficult to apply into computation. [22, Section 6] tries to combine these both ideas together to use the techniques of generalized native spaces to construct RKBSs.

In this paper we want to complete and extend the theoretical results in [22, Section 6]. In addition, the RKBS denoted in Definition 4.1 is different from [17, 18, 23]. The RKBS redefines in weak sense or strong sense and its reproducing kernel K can be set up on nonsymmetric domains, i.e., $K : \Omega_2 \times \Omega_1 \to \mathbb{R}$ where Ω_2 and Ω_1 can be various subsets of $\mathbb{R}^d$ (see Definition 4.1). The RKBS is the extension of the RKHS. The RKBS defined in [17, 18, 23] can be seen as a special case of the RKBS defined in this paper. Even if RKBS is defined in weak sense, we can still obtain the optimal recovery in the weak-sense RKBS using the techniques similar as in [23] (see Lemma 4.1). Next we are able to apply this optimal recovery to solve support vector machines in RKBSs same as the classical problems in RKHSs, i.e., the classical support vector machine in a RKHS $\mathcal{H}$ for the training data $\left\{(\boldsymbol{x}_j, y_j)\right\}_{j=1}^N$ is

$$\min_{f\in\mathcal{H}} \sum_{j=1}^{N} L(\boldsymbol{x}_j, y_j, f(\boldsymbol{x}_j)) + \Sigma\left(\|f\|_{\mathcal{H}}\right),$$

where the loss function L and Σ are given. Suppose that $\mathcal{B}$ is a weak-sense RKBS with a reproducing kernel K defined on $\Omega_2 \times \Omega_1 \subseteq \mathbb{R}^d \times \mathbb{R}^d$. Under some sufficient conditions of the loss function L and Σ, Theorem 4.2 provides that the support vector machine in $\mathcal{B}$ for the training data $\left\{(\boldsymbol{x}_j, y_j)\right\}_{j=1}^N \subset \Omega_1 \times \mathbb{C}$

$$\min_{f\in\mathcal{B}} \sum_{j=1}^{N} L(\boldsymbol{x}_j, y_j, f(\boldsymbol{x}_j)) + \Sigma\left(\|f\|_{\mathcal{B}}\right),$$

has a unique optimal solution $s_{D,L,\Sigma}$ and its dual element is a linear combination of its reproducing kernel centered at the given data points $X = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \subseteq \Omega_1$, i.e.,

$$s^*_{D,L,\Sigma}(\boldsymbol{x}) = \sum_{k=1}^{N} c_k K(\boldsymbol{x}, \boldsymbol{x}_k), \quad \boldsymbol{x} \in \Omega_2.$$

It is obvious that the support vector machine in RKBSs is the generalization form of the classical case in RKHSs.

In Section 4.2, we shows how to use a positive definite function Φ to set up different strong-sense RKBSs $\mathcal{B}^p_\Phi(\mathbb{R}^d)$ and $\mathcal{B}^p_\Phi(\Omega)$ with $p > 1$ whose reproducing kernel is given by $K(\boldsymbol{x}, \boldsymbol{y}) = \Phi(\boldsymbol{x} - \boldsymbol{y})$ (see Theorem 4.3 and 4.7). It is obvious that $\mathcal{B}^p_\Phi(\mathbb{R}^d)$ is a kind of generalized native spaces. Moreover, $\mathcal{B}^p_\Phi(\mathbb{R}^d)$ and $\mathcal{B}^p_\Phi(\Omega)$ coincide with the definition of RKBSs given in [23]. The empirical support vector solution in $\mathcal{B}^p_\Phi(\mathbb{R}^d)$ can be represented by the positive definite function Φ (see Corollary 4.6). Theorem 4.7 gives an example of the nonsymmetric reproducing kernel K defined on $\mathbb{R}^d \times \Omega$. Corollary 4.5

and 4.8 provide that these RKBSs can be further imbedded into some Sobolev spaces for some special reproducing kernels, e.g., Sobolev spline kernels.

The Matérn functions represent a recent and fast growing research area which has frequent applications in approximation theory and statistical learning (see [7, 14]). The papers [9, 22] show that the Matérn functions are positive definite functions and (full-space) Green functions. In Section 5, we use the Matérn functions to solve the support vector machines in their RKBSs. Let $G_{\theta,n}$ be the Matérn function with parameter $\theta > 0$ and degree $n > 3d/2$. According to our theoretical results, $\mathcal{B}^2_{G_{\theta,n}}(\mathbb{R}^d)$ is a RKHS and $\mathcal{B}^4_{G_{\theta,n}}(\mathbb{R}^d)$ is a just RKBS. Their reproducing kernels are given in the same Sobolev spline kernel $K_{\theta,n}(\boldsymbol{x},\boldsymbol{y}) := G_{\theta,n}(\boldsymbol{x}-\boldsymbol{y})$. It is well known that the support vector solution in $\mathcal{B}^2_{G_{\theta,n}}(\mathbb{R}^d) \equiv \mathcal{H}_{G_{\theta,n}}(\mathbb{R}^d)$ has the explicit expression

$$s_{D,L,\Sigma}(\boldsymbol{x}) := \sum_{k=1}^{N} K_{\theta,n}(\boldsymbol{x},\boldsymbol{x}_k), \quad \boldsymbol{x} \in \mathbb{R}^d,$$

(see Theorem 3.1). Here we discover a new fact that the support vector solution in $\mathcal{B}^4_{G_{\theta,n}}(\mathbb{R}^d)$ has the explicit form

$$s_{D,L,\Sigma}(\boldsymbol{x}) = \sum_{j,k,l=1}^{N,N,N} c_j \overline{c_k} c_l K_{\theta,3n}\left(\boldsymbol{x}+\boldsymbol{x}_k, \boldsymbol{x}_j+\boldsymbol{x}_l\right), \quad \boldsymbol{x} \in \mathbb{R}^d,$$

and many other fresh explicit support vector solutions in the RKBS $\mathcal{B}^{2l}_{G_{\theta,n}}(\mathbb{R}^d)$ (see Section 5). This discovery could produce a new numerical tool for the support vector machines in our future research works.

Remark 1.1. In this paper, the author Ye hope to make corrections of his mistake of the optimal recovery of RKBS $\mathcal{B}^p_{\Phi}(\mathbb{R}^d)$ mentioned in [22, Section 6.2]. Corollary 4.6 is the correction of [22, Theorem 6.5], which is misconception on the dual-element map to be linear. But the main ideas and techniques are the same as in his original [22]. The new version of [22] has been updated in Ye's webpage.

2 Banach Spaces

In this paper, we review some classical theoretical results of Banach space in [11, 13, 15, 16]. We denote the dual space (the collection of all bounded linear functionals) of a Banach space $\mathcal{B}$ to be $\mathcal{B}'$ and its dual bilinear product as $\langle\cdot,\cdot\rangle_{\mathcal{B}}$, i.e.,

$$\langle f, T\rangle_{\mathcal{B}} := T(f), \quad \text{for all } T \in \mathcal{B}' \text{ and all } f \in \mathcal{B}.$$

[16, Theorem 1.10.7] that $\mathcal{B}'$ is also a Banach space.

If the Banach spaces $\mathcal{B}_1$ and $\mathcal{B}_2$ are *isometrically isomorphic* (equivalent), i.e., $\mathcal{B}_1 \equiv \mathcal{B}_2$, then we say that the both spaces are identical in the sense of their norms in order that their elements can be seen to be the same in the both spaces (see [16, Definition 1.4.13]). We say that $\mathcal{B}_1$ is *imbedded into* $\mathcal{B}_2$ if there exists a positive constant C such that $\|f\|_{\mathcal{B}_2} \leq C\|f\|_{\mathcal{B}_1}$ for all $f \in \mathcal{B}_1 \subseteq \mathcal{B}_2$ (see [1, Section 1.25]).

If the Banach space $\mathcal{B}$ is *reflexive* (see [16, Definition 1.11.6]), then we have $\mathcal{B}'' \equiv \mathcal{B}$ and $\langle f, g\rangle_{\mathcal{B}} = \langle g, f\rangle_{\mathcal{B}'}$ for all $f \in \mathcal{B}$ and all $g \in \mathcal{B}'$. For example, the function space $\mathrm{L}_p(\Omega;\mu)$ defined on the positive measure space $(\Omega, \mathscr{B}_\Omega, \mu)$ is a reflexive Banach space and its dual space is isometrically equivalent to $\mathrm{L}_q(\Omega;\mu)$ where $p, q \geq 1$ and $p^{-1} + q^{-1} = 1$ (see [16, Example 1.10.2 and Theorem 1.11.10]). For the complex situation, the isometrical isomorphism from $\mathrm{L}_p(\Omega;\mu)'$ onto $\mathrm{L}_q(\Omega;\mu)$ is antilinear.

It is well known that we can discuss the orthogonality in Banach spaces with a more general axiom system than that in Hilbert spaces. The papers [11, 13, 15] show that every Banach space can be represented as an semi-inner-product space in order that the theories of Banach space can be penetrated by Hilbert space type arguments. A *semi-inner product* $[\cdot,\cdot]_{\mathcal{B}} : \mathcal{B}\times\mathcal{B} \to \mathbb{C}$ defined on a Banach space $\mathcal{B}$ is given by

$$
\begin{aligned}
&(i)\ [f+g,h]_{\mathcal{B}} = [f,h]_{\mathcal{B}} + [g,h]_{\mathcal{B}},\\
&(ii)\ [\lambda f,g]_{\mathcal{B}} = \lambda[f,g]_{\mathcal{B}}, \quad [f,\lambda g]_{\mathcal{B}} = \overline{\lambda}[f,g]_{\mathcal{B}},\\
&(iii)\ [f,f]_{\mathcal{B}} = \|f\|_{\mathcal{B}}^2,\\
&(iv)\ |[f,g]_{\mathcal{B}}| \leq [f,f]_{\mathcal{B}}[g,g]_{\mathcal{B}},
\end{aligned}
$$

for all $f,g,h\in\mathcal{B}$ and all $\lambda\in\mathbb{C}$. However, the Heremitian symmetric of the semi-inner product may be not true, i.e., $[f,g]_{\mathcal{B}} \neq \overline{[g,f]_{\mathcal{B}}}$. This indicates that the generality of the semi-inner product of Banach space is a serious limitation on any extensive development parallel to the inner product of Hilbert space.

We call that f is *orthogonal* to g in a Banach space $\mathcal{B}$ if

$$\|f+\lambda g\|_{\mathcal{B}} \geq \|f\|_{\mathcal{B}}, \quad \text{for all } \lambda\in\mathbb{C},$$

(see definitions in [11, 13]). Suppose that the Banach space $\mathcal{B}$ is *smooth* (see [16, Definition 5.4.1]). Using [11, Theorem 2], we can determine that f is orthogonal to g if and only if f is *normal to* g, i.e.,

$$[g,f]_{\mathcal{B}} = 0.$$

According to [11, Theorem 3] and [16, Corollary 5.4.18], the following three statements are equivalent: (i) $\mathcal{B}$ is smooth, (ii) $\mathcal{B}$ is Gâteaux differentiable (see [16, Definition 5.4.15]), (iii) the semi-inner product of $\mathcal{B}$ is continuous (see [11]).

We can also obtain the representation theorem in the Banach space by an adaptation of the representation theorem in the Hilbert space. If the Banach space $\mathcal{B}$ is smooth and *uniformly convex* (see [16, Definition 5.2.1]), then, for every bounded linear functional $T_f\in\mathcal{B}'$, there exists a unique $f\in\mathcal{B}$ such that

$$T_f(g) = [g,f]_{\mathcal{B}}, \quad \text{for all } g\in\mathcal{B},$$

and $\left\|T_f\right\|_{\mathcal{B}'} = \|f\|_{\mathcal{B}}$. We denote T is the *dual element* of f and rewrite it as $f^* := T_f$. The dual-element map is a one-to-one and norm-preserving mapping from $\mathcal{B}'$ onto $\mathcal{B}$. But this map is perhaps nonlinear. [11, Theorem 7] provides that the semi-inner product of $\mathcal{B}'$ has the form $[f^*,g^*]_{\mathcal{B}'} = [g,f]_{\mathcal{B}}$ for all $f^*,g^*\in\mathcal{B}'$. Let $\mathcal{N}$ be a subset of $\mathcal{B}$. We can check that f is orthogonal to $\mathcal{N}$ if and only if its dual element $f^*\in\mathcal{N}^{\perp}$, i.e.,

$$[h,f]_{\mathcal{B}} = \langle h,f^*\rangle_{\mathcal{B}} = 0, \quad \text{for all } h\in\mathcal{N}.$$

For example, the semi-inner product of $\mathrm{L}_p(\Omega;\mu)$ with $1<p<\infty$ is given by

$$[g,f]_{\mathrm{L}_p(\Omega;\mu)} = \frac{1}{\|f\|_{\mathrm{L}_p(\Omega;\mu)}^{p-2}}\int_\Omega g(\boldsymbol{x})\overline{f(\boldsymbol{x})}\,|f(\boldsymbol{x})|^{p-2}\,\mathrm{d}\mu(\boldsymbol{x}), \quad \text{for all } f,g\in\mathrm{L}_p(\Omega;\mu),$$

(see examples in [11, 13]).

3 Reproducing Kernels and Reproducing Kernel Hilbert Spaces

Most of the material presented in this section can be found in the excellent monograph [7, 19, 21]. For the reader's convenience we repeat there what is essential to our discussion later on.

Definition 3.1 ([21, Definition 10.1])**.** Let $\Omega \subseteq \mathbb{R}^d$ and $\mathcal{H}$ be a Hilbert space consisting of functions $f : \Omega \to \mathbb{C}$. $\mathcal{H}$ is called a *reproducing kernel Hilbert space* (RKHS) and a kernel function $K : \Omega \times \Omega \to \mathbb{R}$ is called a *reproducing kernel* for $\mathcal{H}$ if

$$(i)\ K(\cdot, \boldsymbol{y}) \in \mathcal{H} \text{ and } (ii)\ f(\boldsymbol{y}) = (f, K(\cdot, \boldsymbol{y}))_{\mathcal{H}}, \quad \text{for all } f \in \mathcal{H} \text{ and all } \boldsymbol{y} \in \Omega,$$

where $(\cdot, \cdot)_{\mathcal{H}}$ is used to denote the inner product of $\mathcal{H}$.

Remark 3.1. For simplifying the discussion and proof, we let all the kernel function be real-valued and all the function spaces compose of complex-valued functions in this paper. According to [16, Proposition 1.9.3], it is not difficult for us to extend the theoretical results into complex-valued kernel functions or restrict the function spaces into real.

3.1 Optimal Recovery in Reproducing Kernel Hilbert Spaces

Theorem 3.1 (representer theorem [19, Theorem 5.5])**.** *Let $\mathcal{H}$ be a reproducing kernel Hilbert space with a reproducing kernel K defined on $\Omega \subseteq \mathbb{R}^d$ and $\Sigma : [0, \infty) \to [0, \infty)$ be convex and increasing. We choose the loss function $L : \mathbb{R}^d \times \mathbb{C} \times \mathbb{C} \to [0, \infty)$ such that $L(\boldsymbol{x}, y, \cdot)$ is a convex map for any fixed $\boldsymbol{x} \in \mathbb{R}^d$ and any fixed $y \in \mathbb{C}$. Given the pairwise distinct data points $X = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \subseteq \Omega$ and the associated data values $Y = \{y_1, \ldots, y_N\} \subset \mathbb{C}$, the optimal solution (support vector solution) $s_{D,L,\Sigma}$ of*

$$\min_{f \in \mathcal{H}} \sum_{j=1}^{N} L(\boldsymbol{x}_j, y_j, f(\boldsymbol{x}_j)) + \Sigma\left(\|f\|_{\mathcal{H}}\right),$$

has the empirical representation

$$s_{D,L,\Sigma}(\boldsymbol{x}) := \sum_{k=1}^{N} c_k K(\boldsymbol{x}, \boldsymbol{x}_k), \quad \boldsymbol{x} \in \Omega.$$

Here $D := \{(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_N, y_N)\}$.

3.2 Constructing Reproducing Kernel Hilbert Spaces by Positive Definite Functions

Definition 3.2 ([21, Definition 6.1])**.** A continuous even function $\Phi : \mathbb{R}^d \to \mathbb{R}$ is called *positive definite* if, for all $N \in \mathbb{N}$, all sets of pairwise distinct centers $X = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \subset \mathbb{R}^d$, the quadratic form

$$\sum_{j=1}^{N} \sum_{k=1}^{N} c_j \overline{c_k} \Phi(\boldsymbol{x}_j - \boldsymbol{x}_k) = \boldsymbol{c}^* \mathsf{A}_{\Phi,X} \boldsymbol{c} > 0, \quad \text{for all } \boldsymbol{c} \in \mathbb{C}^N \backslash \{\boldsymbol{0}\},$$

where the interpolation matrix $\mathsf{A}_{\Phi,X} := \left(\Phi(\boldsymbol{x}_j - \boldsymbol{x}_k)\right)_{j,k=1}^{N,N} \in \mathbb{C}^{N \times N}$ and $\boldsymbol{c}^* = \overline{\boldsymbol{c}}^T$.

It shows that Φ is positive definite if and only if $\mathsf{A}_{\Phi,X}$ is positive definite for any pairwise distinct finite data points X in $\mathbb{R}^d$. The application and history of positive definite functions can be seen in the review paper [8].

Theorem 3.2 ([21, Theorem 6.11])**.** *Suppose that $\Phi \in \mathrm{C}(\mathbb{R}^d) \cap \mathrm{L}_1(\mathbb{R}^d)$. Then Φ is positive definite if and only if Φ is bounded and its Fourier transform $\hat{\Phi}$ is nonnegative and nonvanishing (nonzero everywhere).*

Remark 3.2. In this paper, the Fourier transform of $f \in \mathrm{L}_1(\mathbb{R}^d)$ is defined by

$$\hat{f}(\boldsymbol{x}) := (2\pi)^{-d/2} \int_{\mathbb{R}^d} f(\boldsymbol{y}) e^{-i\boldsymbol{x}^T\boldsymbol{y}} \mathrm{d}\boldsymbol{y},$$

where i is the imaginary unit, i.e., $i^2 = -1$.

[21, Section 10.2] shows how to use the positive definite functions to construct the RKHSs.

Theorem 3.3 ([21, Theorem 10.12])**.** *Suppose that $\Phi \in \mathrm{C}(\mathbb{R}^d) \cap \mathrm{L}_1(\mathbb{R}^d)$ is a positive definite function. Then the space*

$$\mathcal{H}_\Phi(\mathbb{R}^d) := \left\{ f \in \mathrm{L}_2(\mathbb{R}^d) \cap \mathrm{C}(\mathbb{R}^d) : \hat{f}/\hat{\Phi}^{1/2} \in \mathrm{L}_2(\mathbb{R}^d) \right\},$$

equipped with the norm form

$$\|f\|_{\mathcal{H}_\Phi(\mathbb{R}^d)} := \left((2\pi)^{-d/2} \int_{\mathbb{R}^d} \frac{\left|\hat{f}(\boldsymbol{x})\right|^2}{\hat{\Phi}(\boldsymbol{x})} \mathrm{d}\boldsymbol{x} \right)^{1/2},$$

is a reproducing kernel Hilbert space (native space) with a reproducing kernel given by

$$K(\boldsymbol{x}, \boldsymbol{y}) := \Phi(\boldsymbol{x} - \boldsymbol{y}), \quad \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d,$$

where $\hat{\Phi}$ and $\hat{f}$ are the Fourier transforms of Φ and f. Its inner product has the form

$$(f, g)_{\mathcal{H}} = (2\pi)^{-d/2} \int_{\mathbb{R}} \frac{\hat{f}(\boldsymbol{x})\overline{\hat{g}(\boldsymbol{x})}}{\hat{\Phi}(\boldsymbol{x})} \mathrm{d}\boldsymbol{x}, \quad f, g \in \mathcal{H}_\Phi(\mathbb{R}^d).$$

Using the Fourier transform technique similar as in Theorem 3.3, we can set up RKBSs by the positive definite functions (see Section 4.2).

4 Reproducing Kernels and Reproducing Kernel Banach Spaces

Now we give the definition of RKBSs which is a natural generality of RKHSs by viewing the inner product as the dual bilinear product.

Definition 4.1. Let Ω_1 and Ω_2 be subsets of $\mathbb{R}^d$ and $\mathcal{B}$ be a Banach space composed of functions $f : \Omega_1 \to \mathbb{C}$, whose dual space $\mathcal{B}'$ is isometrically equivalent to a function space $\mathcal{F}$ with $g : \Omega_2 \to \mathbb{C}$. If there is a kernel function $K : \Omega_2 \times \Omega_1 \to \mathbb{R}$ such that

$$(i)\ K(\cdot, \boldsymbol{y}) \in \mathcal{F} \equiv \mathcal{B}' \text{ and } (ii)\ f(\boldsymbol{y}) = \langle f, K(\cdot, \boldsymbol{y}) \rangle_{\mathcal{B}}, \quad \text{for all } f \in \mathcal{B} \text{ and all } \boldsymbol{y} \in \Omega_1,$$

then we call $\mathcal{B}$ a *reproducing kernel Banach space* (RKBS) and K its *reproducing kernel* in *weak sense*.

If the weak-sense RKBS $\mathcal{B}$ is further reflexive and has the other-side reproduction, i.e.,

$$(iii)\ K(\boldsymbol{x},\cdot) \in \mathcal{B} \text{ and } (iv)\ \overline{\langle K(\boldsymbol{x},\cdot), g\rangle_{\mathcal{B}}} = g(\boldsymbol{x}), \quad \text{for all } g \in \mathcal{F} \equiv \mathcal{B}' \text{ and all } \boldsymbol{x} \in \Omega_2,$$

then it is obvious that $\mathcal{F}$ is also a weak-sense RKBS with the RK $\tilde{K}(\boldsymbol{x},\boldsymbol{y}) := K(\boldsymbol{y},\boldsymbol{x})$ defined on $\Omega_1 \times \Omega_2$. With these additional conditions, we say that $\mathcal{B}$ is a *reproducing kernel Banach space* with the *reproducing kernel K* in *strong sense*.

Remark 4.1. We know that the Riesz representer map on complex Hilbert space $\mathcal{H}$ is antilinear, i.e.,

$$T_{\lambda g}(f) = \langle f, \lambda g\rangle_{\mathcal{H}} = \overline{\lambda}(f,g)_{\mathcal{H}} = \overline{\lambda}\langle f, g\rangle_{\mathcal{H}} = \overline{\lambda}T_g(f), \quad \text{for all } f,g \in \mathcal{H} \text{ and all } \lambda \in \mathbb{C}.$$

Here we even let the isometrical isomorphism from the dual space $\mathcal{B}'$ onto the related function space be antilinear. Thus, the format of strong-sense RKBS coincides with the complex RKHS, i.e.,

$$\langle f, K(\cdot,\boldsymbol{y})\rangle_{\mathcal{H}} = \overline{\langle K(\boldsymbol{y},\cdot), f\rangle_{\mathcal{H}}} = (f, K(\cdot,\boldsymbol{y}))_{\mathcal{H}} = f(\boldsymbol{y}), \quad \text{for all } f \in \mathcal{H} \text{ and all } \boldsymbol{y} \in \Omega,$$

which indicates that the RKHS is a special case of strong-sense RKBS.

Why we set up the definition of RKBS different from [23, Definition 1]? The reason is that we can show the optimal recovery in RKBS even if it is defined in weak sense. Moreover, since the dual space of Hilbert space is isometrically equivalent to itself, we can choose the equivalent functions space $\mathcal{F} \equiv \mathcal{H}$ such that the domains of the reproducing kernel K are symmetric, i.e., $\Omega_2 = \Omega_1$. Actually the Banach space $\mathcal{B}$ is usually not equal to any equivalent function space $\mathcal{F}$ of its dual $\mathcal{B}'$ even though we do not require their norms to be equivalent. We naturally do not need the symmetric conditions in the Banach space. So the nonsymmetric domains are used to define the RKBS $\mathcal{B}$ and the reproducing kernel K, i.e., $\Omega_2 \neq \Omega_1$. In this paper, we only consider the nonsymmetric domains chosen in $\mathbb{R}^d$. The domains of K are related to both $\mathcal{B}$ and $\mathcal{F} \equiv \mathcal{B}'$. If we choose different $\mathcal{F}$ to be isometrically equivalent to the dual $\mathcal{B}'$, then we can obtain different reproducing kernel K of the RKBS $\mathcal{B}$ dependent on its equivalent dual space $\mathcal{F}$.

The $K(\cdot,\boldsymbol{y})$ can be seen as a point evaluation function $\delta_{\boldsymbol{y}}$ defined on $\mathcal{B}$. This implies that $\delta_{\boldsymbol{y}}$ is a bounded linear functional on $\mathcal{B}$, i.e., $\delta_{\boldsymbol{y}} \in \mathcal{B}'$. If the Banach space $\mathcal{B}$ is further uniformly convex and smooth, then its semi-inner product and its dual-element map are well-defined. Because the inner product is also the semi-inner product and the dual-element map is an identical map on a Hilbert space. When $\mathcal{B}$ is uniformly convex and smooth, then we can even use the semi-inner product and the dual element map to set up the equivalent conditions of weak-sense RKBSs, i.e.,

$$\delta_{\boldsymbol{y}} \in \mathcal{B}' \equiv \mathcal{F} \text{ which indicates that } f(\boldsymbol{y}) = \langle f, \delta_{\boldsymbol{y}}\rangle_{\mathcal{B}} = [f, \delta_{\boldsymbol{y}}^*]_{\mathcal{B}}, \quad \text{for all } f \in \mathcal{B} \text{ and all } \boldsymbol{y} \in \Omega_1,$$

If $\mathcal{B}$ is a strong-sense RKBS, then the equivalent dual space $\mathcal{F}$ of $\mathcal{B}$ is also a strong-sense RKBS. All RKBSs and reproducing kernels set up in Section 4.2 satisfy the strong-sense definition but their domains can be symmetric or nonsymmetric.

Suppose that a sequence $\{f_n\}_{n=1}^{\infty} \subset \mathcal{B}$ and $f \in \mathcal{B}$ such that $\|f - f_n\|_{\mathcal{B}} \to 0$ when $n \to \infty$. We fix any $\boldsymbol{y} \in \Omega_1$. Then

$$|f(\boldsymbol{y}) - f_n(\boldsymbol{y})| = |\langle f - f_n, K(\cdot,\boldsymbol{y})\rangle_{\mathcal{B}}| \leq \|K(\cdot,\boldsymbol{y})\|_{\mathcal{B}'} \|f - f_n\|_{\mathcal{B}} \to 0,$$

when $n \to \infty$. This means that convergence in the weak-sense RKBS $\mathcal{B}$ implies pointwise convergence.

Let $\mathcal{N}$ be a completion (close) of span $\{K(\cdot, \boldsymbol{y}) : \boldsymbol{y} \in \Omega_1\} \subseteq \mathcal{F} \equiv \mathcal{B}'$ with its dual norm. Now we show that $\mathcal{N} \equiv \mathcal{F} \equiv \mathcal{B}'$ which means that span $\{K(\cdot, \boldsymbol{y}) : \boldsymbol{y} \in \Omega_1\}$ is dense in $\mathcal{F}$ and $\{K(\cdot, \boldsymbol{y}) : \boldsymbol{y} \in \Omega_1\}$ is a linear basis of $\mathcal{F}$ when $\mathcal{B}$ is a reflexive weak-sense RKBS. It is obvious that $\mathcal{N} \subseteq \mathcal{F}$. Assume that $\mathcal{N} \subsetneqq \mathcal{F}$. According to [16, Corollary 1.9.7] there is an element $f \in \mathcal{B} \equiv \mathcal{B}'' \equiv \mathcal{F}'$ such that $\|f\|_{\mathcal{B}} = 1$ and $f(\boldsymbol{y}) = \langle f, K(\cdot, \boldsymbol{y})\rangle_{\mathcal{B}} = 0$ for all $\boldsymbol{y} \in \Omega_1$. We find the contradiction between $\|f\|_{\mathcal{B}} = 1$ and $f = 0$. Thus the first assumption is not true and then we can conclude that $\mathcal{N} \equiv \mathcal{F} \equiv \mathcal{B}'$.

Example 4.1. We give a simple example of the strong-sense RKBS. Let $\Omega_2 = \Omega_1 := \{1, \cdots, n\}$ and $\mathsf{A} \in \mathbb{R}^{n\times n}$ be a symmetric positive definite matrix. It can be decomposed into $\mathsf{A} = \mathsf{V}\mathsf{D}\mathsf{V}^T$, where D is a positive diagonal matrix and V is an orthogonal matrix. We choose $p, q > 1$ such that $p^{-1} + q^{-1} = 1$. Define $\mathcal{B} := \{f : \Omega_1 \to \mathbb{C}\}$ equipped with the norm

$$\|f\|_{\mathcal{B}} := \left\|\mathsf{D}^{-1/q}\mathsf{V}^T \boldsymbol{f}\right\|_q, \quad \text{where } \boldsymbol{f} := (f(1), \cdots, f(n))^T .$$

We can check that $\mathcal{B}$ is a Banach space and its dual space $\mathcal{B}'$ is isometrically equivalent to $\mathcal{F} := \{g : \Omega_2 \to \mathbb{C}\}$ equipped with the norm

$$\|g\|_{\mathcal{B}'} := \left\|\mathsf{D}^{-1/p}\mathsf{V}^T \boldsymbol{g}\right\|_p, \quad \text{where } \boldsymbol{g} := (g(1), \cdots, g(n))^T .$$

Moreover, its dual bilinear form is given by

$$\langle f, g\rangle_{\mathcal{B}} = \boldsymbol{g}^*\mathsf{A}^{-1}\boldsymbol{f}, \quad \text{for all } f \in \mathcal{B} \text{ and all } g \in \mathcal{B}'.$$

If the kernel function is defined by

$$K(j, k) := \mathsf{A}_{jk}, \quad j \in \Omega_2,\ k \in \Omega_1,$$

then the reproduction can be easily verified, i.e.,

$$\langle f, K(\cdot, k)\rangle_{\mathcal{B}} = f(k), \quad k \in \Omega_1, \quad \text{and} \quad \overline{\langle K(j, \cdot), g\rangle_{\mathcal{B}}} = g(j), \quad j \in \Omega_2.$$

Therefore $\mathcal{B}$ is indeed a RKBS in strong sense.

4.1 Optimal Recovery in Reproducing Kernel Banach Spaces

It is well-known that the Hilbert space is uniformly convex and smooth. It is natural for us to assume the RKBS is further uniformly convex and smooth to introduce its optimal recovery in weak-sense RKBS.

Given the pairwise distinct data points $X = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \subseteq \Omega_1$ and the associated data values $Y = \{y_1, \ldots, y_N\} \subset \mathbb{C}$, we define the subset in the RKBS $\mathcal{B}$ by

$$\mathcal{N}_{\mathcal{B}}(X, Y) := \left\{f \in \mathcal{B} : f(\boldsymbol{x}_j) = y_j, \text{ for all } j = 1, \ldots, N\right\}.$$

If $\mathcal{N}_{\mathcal{B}}(X, Y)$ is a null set, then there is no meaning for the support vector machines. So we need to assume that $\mathcal{N}_{\mathcal{B}}(X, Y)$ is always non-null for the given data sites. Actually we can show that $\mathcal{N}_{\mathcal{B}}(X, Y)$ is non-null for any data values Y if and only if $\delta_{\boldsymbol{x}_1}, \ldots, \delta_{\boldsymbol{x}_N}$ are linearly independent on $\mathcal{B}$ because $\sum_{k=1}^N c_k\delta_{\boldsymbol{x}_k} = 0$ if and only if $\sum_{k=1}^N c_k f(\boldsymbol{x}_k) = 0$ for all $f \in \mathcal{B}$ and $\boldsymbol{c} = (c_1, \cdots, c_N)^T = \boldsymbol{0}$ if and only if $\boldsymbol{c}$ orthonormal to $\mathbb{C}^N$. In this section, we assume that $\delta_{\boldsymbol{x}_1}, \ldots, \delta_{\boldsymbol{x}_N}$ are always *linearly independent* on $\mathcal{B}$ for the given data points X, which is equivalent to the fact that $K(\cdot, \boldsymbol{x}_1), \ldots, K(\cdot, \boldsymbol{x}_N)$ are linearly independent.

We use the techniques of [23, Theorem 19] to verify the following lemma.

Lemma 4.1. *Let $\mathcal{B}$ be a weak-sense reproducing kernel Banach space with a reproducing kernel K defined on $\Omega_2 \times \Omega_1 \subseteq \mathbb{R}^d \times \mathbb{R}^d$. Suppose that $\mathcal{B}$ is uniformly convex and smooth. Given the pairwise distinct data points $X = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \subseteq \Omega_1$ and the associated data values $Y = \{y_1, \ldots, y_N\} \subset \mathbb{C}$, the dual element s_D^* of the unique optimal solution s_D of*

$$\min_{f \in \mathcal{B}} \|f\|_{\mathcal{B}} \text{ s.t. } f(\boldsymbol{x}_j) = y_j \text{ for all } j = 1, \ldots, N, \tag{4.1}$$

is the linear combination of $K(\cdot, \boldsymbol{x}_1), \ldots, K(\cdot, \boldsymbol{x}_N)$, i.e.,

$$s_D^*(\boldsymbol{x}) = \sum_{k=1}^{N} c_k K(\boldsymbol{x}, \boldsymbol{x}_k), \quad \boldsymbol{x} \in \Omega_2.$$

Here $D := \{(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_N, y_N)\}$.

Proof. We firstly prove the uniqueness of the optimal solution of the minimum problem (4.1). Let us assume that the minimum problem (4.1) has two optimal solutions $s_1, s_2 \in \mathcal{B}$ with $s_1 \neq s_2$. Since $\mathcal{B}$ is uniformly convex, [16, Corollary] provides that $\left\|\frac{1}{2}(s_1 + s_2)\right\|_{\mathcal{B}} < \frac{1}{2}\|s_1\|_{\mathcal{B}} + \frac{1}{2}\|s_2\|_{\mathcal{B}}$. $\|s_1\|_{\mathcal{B}} = \|s_2\|_{\mathcal{B}}$ then shows for $s_3 := \frac{1}{2}(s_1 + s_2)$ that $\|s_3\|_{\mathcal{B}} < \|s_1\|_{\mathcal{B}}$ and $s_3 \in \mathcal{N}_{\mathcal{B}}(X, Y)$, i.e., s_1 is not a optimal solution of the minimum problem (4.1). The assumption that there are two minimizers is false.

Because $\mathcal{N}_{\mathcal{B}}(X, Y) + \mathcal{N}_{\mathcal{B}}(X, \{0\}) = \mathcal{N}_{\mathcal{B}}(X, Y)$ and $\mathcal{N}_{\mathcal{B}}(X, \{0\})$ is a closed subspace of $\mathcal{B}$. We can determine that the optimal solution s_D is orthogonal to $\mathcal{N}_{\mathcal{B}}(X, \{0\})$, i.e., $\|s_D + h\|_{\mathcal{B}} \geq \|s_D\|_{\mathcal{B}}$ for all $h \in \mathcal{N}_{\mathcal{B}}(X, \{0\})$. Since $\mathcal{B}$ is uniformly convex and smooth, the dual element s_D^* of s_D is well-defined and

$$[h, s_D]_{\mathcal{B}} = \langle h, s_D^* \rangle_{\mathcal{B}} = 0, \text{ for all } h \in \mathcal{N}_{\mathcal{B}}(X, \{0\}),$$

which implies that

$$s_D^* \in \mathcal{N}_{\mathcal{B}}(X, \{0\})^{\perp} = \{g \in \mathcal{F} \equiv \mathcal{B}' : \langle h, g \rangle_{\mathcal{B}} = 0, \text{ for all } h \in \mathcal{N}_{\mathcal{B}}(X, \{0\})\}.$$

It is obvious that

$$\mathcal{N}_{\mathcal{B}}(X, \{0\}) = \left\{f \in \mathcal{B} : f(\boldsymbol{x}_j) = \langle f, K(\cdot, \boldsymbol{x}_j) \rangle_{\mathcal{B}} = 0, \text{ for all } j = 1, \ldots, N\right\} = {}^{\perp}\mathrm{span}\{K(\cdot, \boldsymbol{x}_k)\}_{k=1}^{N}.$$

According to [16, Proposition 1.10.15], we have

$$s_D^* \in \left({}^{\perp}\mathrm{span}\{K(\cdot, \boldsymbol{x}_k)\}_{k=1}^{N}\right)^{\perp} = \mathrm{span}\{K(\cdot, \boldsymbol{x}_1), \ldots, K(\cdot, \boldsymbol{x}_N)\}.$$

□

If $\mathcal{B}$ is uniformly convex and $\Sigma : [0, \infty) \to [0, \infty)$ is convex and increasing, then $\Sigma(\|\cdot\|_{\mathcal{B}})$ is strictly convex on $\mathcal{B}$. Now we verify the representer theorem of the support vector machine in the weak-sense RKBS.

Theorem 4.2. *Let $\mathcal{B}$ be a weak-sense reproducing kernel Banach space with a reproducing kernel K defined on $\Omega_2 \times \Omega_1 \subseteq \mathbb{R}^d \times \mathbb{R}^d$ and $\Sigma : [0, \infty) \to [0, \infty)$ be convex and increasing. Suppose that $\mathcal{B}$ is uniformly convex and smooth. We choose the loss function $L : \mathbb{R}^d \times \mathbb{C} \times \mathbb{C} \to [0, \infty)$ such that $L(\boldsymbol{x}, y, \cdot)$ is a convex map for any fixed $\boldsymbol{x} \in \mathbb{R}^d$ and any fixed $y \in \mathbb{C}$. Given the pairwise distinct data points*

$X = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \subseteq \Omega_1$ and the associated data values $Y = \{y_1, \ldots, y_N\} \subset \mathbb{C}$, the dual element of the unique optimal solution (support vector solution) $s_{D,L,\Sigma}$ of

$$\min_{f \in \mathcal{B}} \sum_{j=1}^{N} L(\boldsymbol{x}_j, y_j, f(\boldsymbol{x}_j)) + \Sigma\left(\|f\|_{\mathcal{B}}\right), \tag{4.2}$$

has the empirical representation

$$s^*_{D,L,\Sigma}(\boldsymbol{x}) := \sum_{k=1}^{N} c_k K(\boldsymbol{x}, \boldsymbol{x}_k), \quad \boldsymbol{x} \in \Omega_2.$$

Here $D := \{(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_N, y_N)\}$.

Proof. Let

$$T_{D,L,\Sigma}(f) := \sum_{j=1}^{N} L(\boldsymbol{x}_j, y_j, f(\boldsymbol{x}_j)) + \Sigma\left(\|f\|_{\mathcal{B}}\right), \quad f \in \mathcal{B}.$$

Since the convergence in $\mathcal{B}$ implies the point wise convergence, we obtain the strict convexity and continuity of $T_{D,L,\Sigma}$ by the strictly convexity of $\Sigma\left(\|\cdot\|_{\mathcal{B}}\right)$ and the convexity of $L\left(\boldsymbol{x}_j, y_j, \cdot\right)$ for all $j = 1, \ldots, N$. The uniformly convex norm indicates its reflexivity by Milman-Pettis Theorem [16, Theorem 5.2.15]. According to the existence of minimizers theorem [19, Theorem A.6.9], $T_{D,L,\Sigma}$ has a global minimum because $\mathcal{B}$ is a reflexive Banach space.

We fix any $f \in \mathcal{B}$. According to Lemma 4.1, there exists an element $s_{f,X}$ whose dual element $s^*_{f,X} \in \operatorname{span}\{K(\cdot, \boldsymbol{x}_k)\}_{k=1}^N$ such that $s_{f,X}$ interpolates the data values $\{f(\boldsymbol{x}_1), \ldots, f(\boldsymbol{x}_N)\}$ at the centers points $X = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$ and $\left\|s_{f,X}\right\|_{\mathcal{B}} \leq \|f\|_{\mathcal{B}}$. This implies that

$$T_{D,L,\Sigma}(s_{f,X}) \leq T_{D,L,\Sigma}(f).$$

Therefore the dual element $s^*_{D,L,\Sigma}$ of the optimal solution $s_{D,L,\Sigma}$ of the minimum problem (4.2) belongs to $\operatorname{span}\{K(\cdot, \boldsymbol{x}_k)\}_{k=1}^N$.

□

Remark 4.2. Since $K(\cdot, \boldsymbol{x}_j)$ can be seen as a point evaluation function $\delta_{\boldsymbol{x}_j}$ defined on $\mathcal{B}$, it indicates that the dual element of $s_{D,L,\Sigma}$ can be also written as a linear combination of $\delta_{\boldsymbol{x}_1}, \ldots, \delta_{\boldsymbol{x}_N}$, i.e., $s^*_{D,L,\Sigma} = \sum_{j=1}^N c_j \delta_{\boldsymbol{x}_j}$.

The uniform convexity and smoothness of $\mathcal{B}$ indicates the uniform convexity and smoothness of its dual $\mathcal{B}' \equiv \mathcal{F}$. If $\mathcal{B}$ is a strong-sense RKBS, then we can further obtain the optimal recovery in $\mathcal{F}$ in the same way, i.e., the dual element of the optimal solution (support vector solution) of

$$\min_{g \in \mathcal{F} \equiv \mathcal{B}'} \sum_{j=1}^{N} L(\boldsymbol{x}_j, y_j, \overline{g(\boldsymbol{x}_j)}) + \Sigma\left(\|g\|_{\mathcal{B}'}\right),$$

is the linear combination of $K(\boldsymbol{x}_1, \cdot), \ldots, K(\boldsymbol{x}_N, \cdot)$, where the data points $X = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \subseteq \Omega_2$.

Moreover, since the dual-element map is an identical map on the Hilbert space and the reproducing kernel of RKHS is symmetric, the optimal recovery in RKBS as in Theorem 4.2 can be seen as the generalization form in RKHS as in Theorem 3.1.

4.2 Constructing Reproducing Kernel Banach Spaces by Positive Definite Functions

Now we construct the RKBS by the positive definite function in a similar way of RKHS as Theorem 3.3. Let $1 < p, q < \infty$ and $p^{-1} + q^{-1} = 1$. Suppose that $\Phi \in \mathrm{C}(\mathbb{R}^d) \cap \mathrm{L}_1(\mathbb{R}^d)$ is a positive definite function. According to Theorem 3.2, we know that $\hat{\Phi} \in \mathrm{L}_1(\mathbb{R}^d) \cap \mathrm{C}(\mathbb{R}^d)$ is nonnegative and nonvanishing. We define

$$\begin{aligned}\mathcal{B}_\Phi^p(\mathbb{R}^d) := \Big\{ f \in \mathrm{C}(\mathbb{R}^d) \cap \mathcal{SI} :\ &\text{the distributional Fourier transform } \hat{f} \text{ of } f \\ &\text{is a measurable function defined on } \mathbb{R}^d \text{ such that } \hat{f}/\hat{\Phi}^{1/q} \in \mathrm{L}_q(\mathbb{R}^d) \Big\},\end{aligned}$$

equipped with the norm

$$\|f\|_{\mathcal{B}_\Phi^p(\mathbb{R}^d)} := \left((2\pi)^{-d/2} \int_{\mathbb{R}^d} \frac{\left|\hat{f}(\boldsymbol{x})\right|^q}{\hat{\Phi}(\boldsymbol{x})} \mathrm{d}\boldsymbol{x} \right)^{1/q},$$

where $\mathcal{SI}$ is the collection of all slowly increasing functions (see [21, Definition 5.19]). We define $\mathcal{B}_\Phi^q(\mathbb{R}^d)$ in an analogous way as above.

Remark 4.3. Following the theoretical results of [12, Section 7.1] we can define the distributional Fourier transform $\hat{T} \in \mathscr{S}'$ of the tempered distribution $T \in \mathscr{S}'$ by

$$\langle \gamma, \hat{T} \rangle_{\mathscr{S}} := \langle \hat{\gamma}, T \rangle_{\mathscr{S}}, \quad \text{for all } \gamma \in \mathscr{S},$$

where $\mathscr{S}$ is the Schwartz space (see [21, Definition 5.17]) and $\mathscr{S}'$ is its dual space with the dual bilinear form $\langle \cdot, \cdot \rangle_{\mathscr{S}}$. We can also verify that $\mathrm{C}(\mathbb{R}^d) \cap \mathcal{SI} \subset \mathrm{L}_1^{loc}(\mathbb{R}^d) \cap \mathcal{SI}$ is embedded into $\mathscr{S}'$.

Suppose that $q \leq p$. Since $\hat{\Phi} \in \mathrm{L}_1(\mathbb{R}^d) \cap \mathrm{C}(\mathbb{R}^d)$ and $p/q \geq 1$, $\hat{\Phi}^{p/q} \in \mathrm{L}_1(\mathbb{R}^d)$ which will be used in the proof of the following theorem. One additional symmetric condition of $\hat{\Phi}^{q/p} \in \mathrm{L}_1(\mathbb{R}^d)$ is also needed in the proof.

The positive measure μ on $\mathbb{R}^d$ is well-defined by

$$\mu(A) := (2\pi)^{-d/2} \int_A \frac{\mathrm{d}\boldsymbol{x}}{\hat{\Phi}(\boldsymbol{x})}, \quad \text{for any open set } A \text{ of } \mathbb{R}^d.$$

So the space $\mathrm{L}_q(\mathbb{R}^d;\mu)$ is well-defined on the positive measure space $(\mathbb{R}^d, \mathscr{B}_{\mathbb{R}^d}, \mu)$, i.e.,

$$\mathrm{L}_q(\mathbb{R}^d;\mu) := \left\{ f : \mathbb{R}^d \to \mathbb{C} :\ f \text{ is measurable and } \int_{\mathbb{R}^d} |f(\boldsymbol{x})|^q \,\mathrm{d}\mu(\boldsymbol{x}) < \infty \right\},$$

equipped with the norm

$$\|f\|_{\mathrm{L}_q(\mathbb{R}^d;\mu)} := \left(\int_{\mathbb{R}^d} |f(\boldsymbol{x})|^q \,\mathrm{d}\mu(\boldsymbol{x}) \right)^{1/q}.$$

$\mathrm{L}_p(\mathbb{R}^d;\mu)$ is defined in an analogous way. It is well-known that $\mathrm{L}_q(\mathbb{R}^d;\mu)$ is a Banach space and its dual space $\mathrm{L}_q(\mathbb{R}^d;\mu)'$ is isometrically equivalent to $\mathrm{L}_p(\mathbb{R}^d;\mu)$. In the same way of representation theorem on Hilbert space, the bounded linear functional $T_g \in \mathrm{L}_q(\mathbb{R}^d;\mu)'$ associated with $g \in \mathrm{L}_p(\mathbb{R}^d;\mu)$ is defined by

$$T_g(f) := \int_{\mathbb{R}^d} f(\boldsymbol{x})\overline{g(\boldsymbol{x})}\mathrm{d}\mu(\boldsymbol{x}), \quad \text{for all } f \in \mathrm{L}_q(\mathbb{R}^d;\mu).$$

Here, this isometrical isomorphism from $\mathrm{L}_q(\mathbb{R}^d;\mu)'$ onto $\mathrm{L}_p(\mathbb{R}^d;\mu)$ is antilinear same as the dual of complex Hilbert spaces, i.e.,

$$T_{\lambda g}(f) = \int_{\mathbb{R}^d} f(\boldsymbol{x})\overline{\lambda g(\boldsymbol{x})}\mathrm{d}\mu(\boldsymbol{x}) = \overline{\lambda}T_g(f), \quad \text{for all } f \in \mathrm{L}_q(\mathbb{R}^d;\mu) \text{ and all } \lambda \in \mathbb{C}.$$

If we can show that $\mathcal{B}_\Phi^p(\mathbb{R}^d)$ and $\mathrm{L}_q(\mathbb{R}^d;\mu)$ are isometrically isomorphic, then $\mathcal{B}_\Phi^p(\mathbb{R}^d)$ is a reflexive Banach space and its dual space $\mathcal{B}_\Phi^p(\mathbb{R}^d)'$ is isometrically equivalent to $\mathrm{L}_p(\mathbb{R}^d;\mu)$. It is analogous to $\mathcal{B}_\Phi^q(\mathbb{R}^d) \equiv \mathrm{L}_p(\mathbb{R}^d;\mu)$. If we can further check the both sides reproduction of $\mathcal{B}_\Phi^p(\mathbb{R}^d)$, then $\mathcal{B}_\Phi^p(\mathbb{R}^d)$ is a strong-sense RKBS.

Theorem 4.3. *Let $1 < q \leq 2 \leq p < \infty$ and $p^{-1} + q^{-1} = 1$. Suppose that $\Phi \in \mathrm{L}_1(\mathbb{R}^d) \cap \mathrm{C}(\mathbb{R}^d)$ is a positive definite function on $\mathbb{R}^d$ and that $\hat{\Phi}^{q/p} \in \mathrm{L}_1(\mathbb{R}^d)$. Then $\mathcal{B}_\Phi^p(\mathbb{R}^d)$ is a strong-sense reproducing kernel Banach space with a reproducing kernel*

$$K(\boldsymbol{x},\boldsymbol{y}) := \Phi(\boldsymbol{x}-\boldsymbol{y}), \quad \boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^d.$$

Its dual space $\mathcal{B}_\Phi^p(\mathbb{R}^d)'$ and $\mathcal{B}_\Phi^q(\mathbb{R}^d)$ are isometrically isomorphic. Moreover, $\mathcal{B}_\Phi^p(\mathbb{R}^d)$ is uniformly convex and smooth. In particular, when $p = 2$ then $\mathcal{B}_\Phi^2(\mathbb{R}^d) = \mathcal{H}_\Phi(\mathbb{R}^d)$ is a reproducing kernel Hilbert space same as in Theorem 3.3.

Proof. We firstly prove that $\mathcal{B}_\Phi^p(\mathbb{R}^d)$ and $\mathrm{L}_q(\mathbb{R}^d;\mu)$ are isometrically isomorphic. The Fourier transform map can be seen as a one-to-one map from $\mathcal{B}_\Phi^p(\mathbb{R}^d)$ into $\mathrm{L}_q(\mathbb{R}^d;\mu)$. We can check the identity of their norm

$$\|f\|_{\mathcal{B}_\Phi^p(\mathbb{R}^d)} = \left((2\pi)^{-d/2}\int_{\mathbb{R}^d}\frac{\left|\hat{f}(\boldsymbol{x})\right|^q}{\hat{\Phi}(\boldsymbol{x})}\mathrm{d}\boldsymbol{x}\right)^{1/q} = \left(\int_{\mathbb{R}^d}\left|\hat{f}(\boldsymbol{x})\right|^q\mathrm{d}\mu(\boldsymbol{x})\right)^{1/q} = \left\|\hat{f}\right\|_{\mathrm{L}_q(\mathbb{R}^d;\mu)}.$$

So the Fourier transform map is an isometric isomorphism. Now we prove that the Fourier transform map is surjective. Fix any $h \in \mathrm{L}_q(\mathbb{R}^d;\mu)$. We want to find an element in $\mathcal{B}_\Phi^p(\mathbb{R}^d)$ whose Fourier transform is equal to h. We conclude that $h \in \mathrm{L}_1(\mathbb{R}^d)$ because

$$\int_{\mathbb{R}^d}|h(\boldsymbol{x})|\,\mathrm{d}\boldsymbol{x} \leq \left(\int_{\mathbb{R}^d}\frac{|h(\boldsymbol{x})|^q}{\hat{\Phi}(\boldsymbol{x})}\mathrm{d}\boldsymbol{x}\right)^{1/q}\left(\int_{\mathbb{R}^d}\hat{\Phi}(\boldsymbol{x})^{p/q}\mathrm{d}\boldsymbol{x}\right)^{1/p} < \infty.$$

Thus, the inverse Fourier transform of h as $\check{h}(\boldsymbol{x}) = (2\pi)^{-d/2}\int_{\mathbb{R}^d}h(\boldsymbol{y})e^{i\boldsymbol{x}^T\boldsymbol{y}}\mathrm{d}\boldsymbol{y}$ is well-defined and an element of $\mathrm{C}(\mathbb{R}^d) \cap \mathcal{SI}$. This shows that $\check{h} \in \mathcal{B}_\Phi^p(\mathbb{R}^d)$ and $\hat{\check{h}} = h$ because $\langle\hat{\check{h}},\gamma\rangle_{\mathscr{S}} = \langle h,\check{\hat{\gamma}}\rangle_{\mathscr{S}} = \langle h,\gamma\rangle_{\mathscr{S}}$ for all $\gamma \in \mathscr{S}$. Therefore $\mathcal{B}_\Phi^p(\mathbb{R}^d)$ is isometrically equivalent to $\mathrm{L}_q(\mathbb{R}^d;\mu)$.

Using $\hat{\Phi}^{q/p} \in \mathrm{L}_1(\mathbb{R}^d)$ we can also prove that $\mathcal{B}_\Phi^q(\mathbb{R}^d) \equiv \mathrm{L}_p(\mathbb{R}^d;\mu)$ in an analogous way. Therefore $\mathcal{B}_\Phi^q(\mathbb{R}^d)$ is isometrically equivalent to the dual space of $\mathcal{B}_\Phi^p(\mathbb{R}^d)$.

We fix any $\boldsymbol{y} \in \mathbb{R}^d$. The Fourier transform of $K(\cdot,\boldsymbol{y})$ is equal to $\hat{k}_{\boldsymbol{y}}(\boldsymbol{x}) := \hat{\Phi}(\boldsymbol{x})e^{-i\boldsymbol{x}^T\boldsymbol{y}}$. Since $\hat{\Phi}^{p/q} \in \mathrm{L}_1(\mathbb{R}^d)$ we have $\hat{k}_{\boldsymbol{y}} \in \mathrm{L}_p(\mathbb{R}^d;\mu)$. Thus $K(\cdot,\boldsymbol{y})$ can be seen as an element of $\mathcal{B}_\Phi^q(\mathbb{R}^d) \equiv \mathcal{B}_\Phi^p(\mathbb{R}^d)'$. In addition, we also have $K(\boldsymbol{x},\cdot) \in \mathcal{B}_\Phi^p(\mathbb{R}^d)$ because $\hat{k}_{\boldsymbol{x}} \in \mathrm{L}_q(\mathbb{R}^d;\mu)$ for any $\boldsymbol{x} \in \mathbb{R}^d$.

Finally we verify the reproduction. Fix any $f \in \mathcal{B}_\Phi^p(\mathbb{R}^d)$ and $\boldsymbol{y} \in \mathbb{R}^d$. We can verify that $\hat{f} \in \mathrm{L}_1(\mathbb{R}^d)$ as in the above proof. Moreover, the continuity of f and $\check{\hat{f}}$ allows us to recover f pointwise from its Fourier transform via

$$f(\boldsymbol{x}) = \check{\hat{f}}(\boldsymbol{x}) := (2\pi)^{-d/2}\int_{\mathbb{R}^d}\hat{f}(\boldsymbol{y})e^{i\boldsymbol{x}^T\boldsymbol{y}}\mathrm{d}\boldsymbol{y}.$$

Thus, we have

$$\begin{aligned}\langle f, K(\cdot,\boldsymbol{y})\rangle_{\mathcal{B}^p_\Phi(\mathbb{R}^d)} = \langle \hat{f}, \hat{k}_{\boldsymbol{y}}\rangle_{\mathrm{L}_q(\mathbb{R}^d;\mu)} &= \int_{\mathbb{R}^d} \hat{f}(\boldsymbol{x})\overline{\hat{k}_{\boldsymbol{y}}(\boldsymbol{x})}\mathrm{d}\mu(\boldsymbol{x})\\ &= (2\pi)^{-d/2}\int_{\mathbb{R}^d}\frac{\hat{f}(\boldsymbol{x})\overline{\hat{\Phi}(\boldsymbol{x})e^{-i\boldsymbol{x}^T\boldsymbol{y}}}}{\hat{\Phi}(\boldsymbol{x})}\mathrm{d}\boldsymbol{x}\\ &= (2\pi)^{-d/2}\int_{\mathbb{R}^d}\hat{f}(\boldsymbol{x})e^{i\boldsymbol{x}^T\boldsymbol{y}}\mathrm{d}\boldsymbol{x} = f(\boldsymbol{y}).\end{aligned}$$

In the same way, we can also verify that $\mathcal{B}^p_\Phi(\mathbb{R}^d)$ has the other-side reproduction, i.e.,

$$\overline{\langle K(\boldsymbol{x},\cdot), g\rangle}_{\mathcal{B}^p_\Phi(\mathbb{R}^d)} = \overline{\langle \hat{k}_{\boldsymbol{x}}, \hat{g}\rangle}_{\mathrm{L}_q(\mathbb{R}^d;\mu)} = g(\boldsymbol{x}), \quad \text{for all } g\in\mathcal{B}^q_\Phi(\mathbb{R}^d)\equiv\mathcal{B}^p_\Phi(\mathbb{R}^d)' \text{ and all } \boldsymbol{x}\in\mathbb{R}^d.$$

Therefore $\mathcal{B}^p_\Phi(\mathbb{R}^d)$ is a strong-sense RKBS.

Since $\mathrm{L}_q(\mathbb{R}^d;\mu)$ is uniformly convex and smooth by [16, Theorem 5.2.11 and Example 5.4.8], $\mathcal{B}^p_\Phi(\mathbb{R}^d)\equiv\mathrm{L}_q(\mathbb{R}^d;\mu)$ is also uniformly convex and smooth. □

Remark 4.4. It is obvious that $\mathcal{B}^q_\Phi(\mathbb{R}^d)$ is a also a RKBS with the reproducing kernel K. Combining with [16, Proposition 1.9.3], the restriction of $\mathcal{B}^p_\Phi(\mathbb{R}^d)$ in real is also a RKBS with the reproducing kernel K and its dual is isometrically equivalent to the restriction of $\mathcal{B}^q_\Phi(\mathbb{R}^d)$ in real. It is well-known that the RKHS of the given reproducing kernel is unique. But Theorem 4.3 shows that the reproducing kernels of different RKBSs can be identical.

Corollary 4.4. *Let $\mathcal{B}^p_\Phi(\mathbb{R}^d)$ with $p\geq 2$ be defined in Theorem 4.3. Then $\mathcal{B}^p_\Phi(\mathbb{R}^d)\subseteq\mathrm{L}_p(\mathbb{R}^d)$.*

Proof. We fix any $f\in\mathcal{B}^p_\Phi(\mathbb{R}^d)$. According to the proof of Theorem 4.3, we have $\hat{f}\in\mathrm{L}_q(\mathbb{R}^d)$ because

$$\int_{\mathbb{R}^d}\left|\hat{f}(\boldsymbol{x})\right|^q\mathrm{d}\boldsymbol{x}\leq(2\pi)^{qd/2}\left(\int_{\mathbb{R}^d}\frac{\left|\hat{f}(\boldsymbol{x})\right|^q}{\hat{\Phi}(\boldsymbol{x})}\mathrm{d}\boldsymbol{x}\right)\left(\sup_{\boldsymbol{x}\in\mathbb{R}^d}\hat{\Phi}(\boldsymbol{x})\right)<\infty.$$

The Hausdorff-Young inequality [12, Theorem 7.1.13] provides that $f=\check{\hat{f}}\in\mathrm{L}_p(\mathbb{R}^d)$ because $1<q\leq 2$. □

Remark 4.5. The reproducing kernel Banach space $\mathcal{B}^p_\Phi(\mathbb{R}^d)$ can be precisely written as

$$\begin{aligned}\mathcal{B}^p_\Phi(\mathbb{R}^d) := \Big\{ f\in\mathrm{L}_p(\mathbb{R}^d)\cap\mathrm{C}(\mathbb{R}^d): &\text{ the distributional Fourier transform } \hat{f} \text{ of } f\\ &\text{is a measurable function defined on } \mathbb{R}^d \text{ such that } \hat{f}/\hat{\Phi}^{1/q}\in\mathrm{L}_q(\mathbb{R}^d)\Big\}.\end{aligned}$$

However, $\mathcal{B}^q_\Phi(\mathbb{R}^d)\nsubseteq\mathrm{L}_q(\mathbb{R}^d)$ because the Hausdorff-Young inequality does not work for $p>2$.

We fix any positive number $m>d/2$. According to [21, Corollary 10.13], if there are two positive constants C_1, C_2 such that

$$C_1\left(1+\|\boldsymbol{x}\|_2^2\right)^{-m/2}\leq\hat{\Phi}(\boldsymbol{x})^{1/2}\leq C_2\left(1+\|\boldsymbol{x}\|_2^2\right)^{-m/2}, \quad \boldsymbol{x}\in\mathbb{R}^d,$$

then the RKHS $\mathcal{B}^2_\Phi(\mathbb{R}^d)\equiv\mathcal{H}_\Phi(\mathbb{R}^d)$ and the classical L_2-based Sobolev space $W^m_2(\mathbb{R}^d)\equiv\mathcal{H}^m(\mathbb{R}^d)$ of order m are isomorphic, i.e., $\mathcal{H}_\Phi(\mathbb{R}^d)\cong\mathcal{H}^m(\mathbb{R}^d)$.

Following the ideas of RKHS, we can also find the relationship between the RKBS and the Sobolev spaces. Let $f_m(\boldsymbol{x}) := \left(1+\|\boldsymbol{x}\|_2^2\right)^{m/2}\hat{f}(\boldsymbol{x})$ with $p\geq 2$. The theory of singular integrals then shows that f

belongs to the classical L_p-based Sobolev space $W_p^m(\mathbb{R}^d)$ of order m if any only if the function f_m is the Fourier transform of some function in $\mathrm{L}_p(\mathbb{R}^d)$, and the L_p-norm of the inverse Fourier transform $\check{f}_m$ is equivalent to the W_p^m-norm of f (much more detail mentioned in [1, Section 7.63] and [12, Section 7.9]). Using the Hausdorff-Young inequality, we can set out $\|f\|_{W_p^m(\mathbb{R}^d)} \le C \left\|\check{f}_m\right\|_{\mathrm{L}_p(\mathbb{R}^d)} \le C \|f_m\|_{\mathrm{L}_q(\mathbb{R}^d)}$ for some positive constant C independent on f. Following these statements, we can introduce the following corollary.

Corollary 4.5. *Let $\mathcal{B}_\Phi^p(\mathbb{R}^d)$ with $p \ge 2$ be defined as in Theorem 4.3 and $W_p^m(\mathbb{R}^d)$ be the classical L_p-based Sobolev space of order $m > pd/q - d/q$. Here q is the conjugate exponent of p. If there are two positive constants C_1, C_2 such that*

$$C_1 \left(1 + \|\boldsymbol{x}\|_2^2\right)^{-m/2} \le \hat{\Phi}(\boldsymbol{x})^{1/q} \le C_2 \left(1 + \|\boldsymbol{x}\|_2^2\right)^{-m/2}, \quad \boldsymbol{x} \in \mathbb{R}^d,$$

then $\mathcal{B}_\Phi^p(\mathbb{R}^d)$ is imbedded into $W_p^m(\mathbb{R}^d)$, i.e.,

$$\|f\|_{W_p^m(\mathbb{R}^d)} \le C \|f\|_{\mathcal{B}_\Phi^p(\mathbb{R}^d)}, \quad f \in \mathcal{B}_\Phi^p(\mathbb{R}^d) \subseteq W_p^m(\mathbb{R}^d),$$

for some positive constant C independent on f.

Remark 4.6. Here the lower bound of m is induced by the condition of $\hat{\Phi}^{q/p} \in \mathrm{L}_1(\mathbb{R}^d)$. According to Corollary 4.5, the dual space $W_q^{-m}(\mathbb{R}^d)$ of the Sobolev space $W_p^m(\mathbb{R}^d)$ is imbedded into the dual space $\mathcal{B}_\Phi^p(\mathbb{R}^d)'$ of the RKBS $\mathcal{B}_\Phi^p(\mathbb{R}^d)$. It is well-known that the point evaluation function $\delta_{\boldsymbol{x}}$ belongs to $W_q^{-m}(\mathbb{R}^d)$ (see [1, Section 3.25]) which coincides with $\delta_{\boldsymbol{x}} \in \mathcal{B}_\Phi^p(\mathbb{R}^d)'$.

Since $K(\cdot, \boldsymbol{x}_1), \ldots, K(\cdot, \boldsymbol{x}_N)$ are linear independent in $\mathcal{B}_\Phi^q(\mathbb{R}^d) \equiv \mathcal{B}_\Phi^p(\mathbb{R}^d)'$ for any pairwise distinct data points $X = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \subseteq \mathbb{R}^d$, $\delta_{\boldsymbol{x}_1}, \ldots, \delta_{\boldsymbol{x}_N}$ are linearly independent on $\mathcal{B}_\Phi^p(\mathbb{R}^d)$. According to Theorem 4.2, we can solve the support vector machines in $\mathcal{B}_\Phi^p(\mathbb{R}^d)$ with $p > 1$. Since the dual element of $s_{D,L,\Sigma}^*$ is equal to $s_{D,L,\Sigma}$, we can also obtain the empirical support vector solution.

Corollary 4.6. *Let $\mathcal{B}_\Phi^p(\mathbb{R}^d)$ with $p > 1$ be defined as in Theorem 4.3 and $\Sigma : [0, \infty) \to [0, \infty)$ be convex and increasing. We choose the loss function $L : \mathbb{R}^d \times \mathbb{C} \times \mathbb{C} \to [0, \infty)$ such that $L(\boldsymbol{x}, y, \cdot)$ is a convex map for any fixed $\boldsymbol{x} \in \mathbb{R}^d$ and any fixed $y \in \mathbb{C}$. Given the pairwise distinct data points $X = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \subseteq \mathbb{R}^d$ and the associated data values $Y = \{y_1, \ldots, y_N\} \subset \mathbb{C}$, the unique optimal solution (support vector solution) $s_{D,L,\Sigma}$ of*

$$\min_{f \in \mathcal{B}_\Phi^p(\mathbb{R}^d)} \sum_{j=1}^N L(\boldsymbol{x}_j, y_j, f(\boldsymbol{x}_j)) + \Sigma\left(\|f\|_{\mathcal{B}_\Phi^p(\mathbb{R}^d)}\right),$$

has the empirical representation

$$s_{D,L,\Sigma}(\boldsymbol{x}) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} \hat{\Phi}(\boldsymbol{y})^{p/q} \sum_{k=1}^N c_k e^{i(\boldsymbol{x}-\boldsymbol{x}_k)^T \boldsymbol{y}} \left| \sum_{k=1}^N c_k e^{-i\boldsymbol{x}_k^T \boldsymbol{y}} \right|^{p-2} \mathrm{d}\boldsymbol{y}, \quad \boldsymbol{x} \in \mathbb{R}^d,$$

where q is the conjugate exponent of p and $D := \{(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_N, y_N)\}$.

Proof. We know that the dual element of $f \in \mathrm{L}_p(\mathbb{R}^d; \mu) \equiv \mathrm{L}_q(\mathbb{R}^d; \mu)'$ is given by

$$f^*(\boldsymbol{x}) = \frac{f(\boldsymbol{x}) |f(\boldsymbol{x})|^{p-2}}{\|f\|_{\mathrm{L}_p(\mathbb{R}^d;\mu)}^{p-2}}, \quad \boldsymbol{x} \in \mathbb{R}^d.$$

Using the Fourier transform and the inverse Fourier transform, we can get the empirical form of the optimal solution $s_{D,L,\Sigma}$. □

Remark 4.7. In particular, if p is an even positive integer, then $s_{D,L,\Sigma}$ is also a linear combination of some kernel function transposed at the data points X. For example, when $p = 4$, then

$$s_{D,L,\Sigma}(\boldsymbol{x}) = \sum_{j,k,l=1}^{N,N,N} c_j\overline{c_k}c_l\Phi_3\left(\boldsymbol{x}-\boldsymbol{x}_j+\boldsymbol{x}_k-\boldsymbol{x}_l\right) = \sum_{j,k,l=1}^{N,N,N} c_j\overline{c_k}c_lK_3\left(\boldsymbol{x}+\boldsymbol{x}_k,\boldsymbol{x}_j+\boldsymbol{x}_l\right), \quad \boldsymbol{x}\in\mathbb{R}^d,$$

where the kernel function $K_3(\boldsymbol{x},\boldsymbol{y}) := \Phi_3(\boldsymbol{x}-\boldsymbol{y})$ and Φ_3 is the inverse Fourier transform of $\hat{\Phi}^3$. Moereover,

$$\begin{aligned}&\left\|s_{D,L,\Sigma}\right\|^2_{\mathcal{B}^p_\Phi(\mathbb{R}^d)} = [s_{D,L,\Sigma}, s_{D,L,\Sigma}]_{\mathcal{B}^p_\Phi(\mathbb{R}^d)} = \langle s_{D,L,\Sigma}, s^*_{D,L,\Sigma}\rangle_{\mathcal{B}^p_\Phi(\mathbb{R}^d)}\\ =& \sum_{j=1}^{N}\overline{c_j}\langle s_{D,L,\Sigma}, K(\cdot,\boldsymbol{x}_j)\rangle_{\mathcal{B}^p_\Phi(\mathbb{R}^d)} = \sum_{j,k,l,n=1}^{N,N,N,N}\overline{c_j}c_k\overline{c_l}c_nK_3\left(\boldsymbol{x}_j+\boldsymbol{x}_k,\boldsymbol{x}_l+\boldsymbol{x}_n\right).\end{aligned}$$

In this section, we can think that the RKBS can be given in the p-norm, where $1 < p < \infty$. The generalized Sobolev spaces in [6, 5] can be further set up with the 1-norm. But it is well-known that the 1-norm space is non-uniform-convex, non-smooth or even non-reflexive. So we can not define a 1-norm RKBS in strong sense which indicates that the 1-norm RKBS does not fit the definition of RKBSs given in [23] also. Moreover, the optimal recovery of the support vector solutions may be not true for the 1-norm cases because there may be no semi-inner products induced in the 1-norm RKBSs. The papers [17, 18] show that we can solve the support vector machines in some special 1-norm RKBSs with the typical reproducing kernels. We will try to use the idea of the weak-sense RKBSs to construct the 1-norm RKBSs and solve the support vector machines for the general 1-norm cases in our future research.

We use the techniques of [3, Theorem 6] to set up a strong-sense RKBS defined on a subset Ω of $\mathbb{R}^d$.

Theorem 4.7. *Let $\mathcal{B}^p_\Phi(\mathbb{R}^d)$ with $p > 1$ be defined as in Theorem 4.3 and $\Omega \subseteq \mathbb{R}^d$. Then the function space*

$$\mathcal{B}^p_\Phi(\Omega) := \left\{h : \text{ there exists a function } h \in \mathcal{B}^p_\Phi(\mathbb{R}^d) \text{ such that } f|_\Omega = h\right\},$$

equipped with the norm

$$\|h\|_{\mathcal{B}^p_\Phi(\Omega)} := \min_{h\in\mathcal{B}^p_\Phi(\mathbb{R}^d)} \|f\|_{\mathcal{B}^p_\Phi(\Omega)} \text{ s.t. } f|_\Omega = h,$$

is a strong-sense reproducing kernel Banach space with the reproducing kernel

$$K(\boldsymbol{x},\boldsymbol{y}) := \Phi(\boldsymbol{x}-\boldsymbol{y}), \quad \boldsymbol{x}\in\mathbb{R}^d,\ \boldsymbol{y}\in\Omega,$$

where $f|_\Omega$ stands for the restriction of f into Ω. Its dual space $\mathcal{B}^p_\Phi(\Omega)'$ is isometrically equivalent to a closed subspace of $\mathcal{B}^q_\Phi(\mathbb{R}^d)$ (the annihilator of $\mathcal{N}_0$ in $\mathcal{B}^q_\Phi(\mathbb{R}^d)$)

$$\mathcal{N}_0^\perp = \left\{g\in\mathcal{B}^q_\Phi(\mathbb{R}^d)\equiv\mathcal{B}^p_\Phi(\mathbb{R}^d)' : \langle f,g\rangle_{\mathcal{B}^p_\Phi(\mathbb{R}^d)} = 0, \text{ for all } f\in\mathcal{N}_0\right\},$$

where q is the conjugate exponent of p and

$$\mathcal{N}_0 := \left\{f\in\mathcal{B}^p_\Phi(\mathbb{R}^d) : f|_\Omega = 0\right\}.$$

Moreover, $\mathcal{B}^p_\Phi(\Omega)$ is uniformly convex and smooth.

Proof. Since the convergence in the RKBS $\mathcal{B}_\Phi^p(\mathbb{R}^d)$ implies that the point wise convergence, we can determine that $\mathcal{N}_0$ is a closed subspace of $\mathcal{B}_\Phi^p(\mathbb{R}^d)$. According to the construction of $\mathcal{B}_\Phi^p(\Omega)$, $\mathcal{B}_\Phi^p(\Omega)$ is isometrically equivalent to the quotient space $\mathcal{B}_\Phi^p(\mathbb{R}^d)/\mathcal{N}_0$ (see [16, Definition 1.7.1 and 1.7.3]). Thus $\mathcal{B}_\Phi^p(\Omega)$ is a reflexive Banach space by [16, Theorem 1.7.9 and Corollary 1.11.19].

Next we use the identification of $\left(\mathcal{B}_\Phi^p(\mathbb{R}^d)/\mathcal{N}_0\right)' \equiv \mathcal{N}_0^\perp$ to verify the reproduction (see [16, Theorem 1.10.17]). We fix any $\boldsymbol{y} \in \Omega$. Since

$$\langle f, K(\cdot,\boldsymbol{y})\rangle_{\mathcal{B}_\Phi^p(\mathbb{R}^d)} = f(\boldsymbol{y}) = 0, \quad \text{for all } f \in \mathcal{N}_0,$$

we have $K(\cdot,\boldsymbol{y}) \in \mathcal{N}_0^\perp \equiv \left(\mathcal{B}_\Phi^p(\mathbb{R}^d)/\mathcal{N}_0\right)' \equiv \mathcal{B}_\Phi^p(\Omega)'$. Combining this with the reproduction of $\mathcal{B}_\Phi^p(\mathbb{R}^d)$, we have

$$\langle h, K(\cdot,\boldsymbol{y})\rangle_{\mathcal{B}_\Phi^p(\Omega)} = \langle Eh, K(\cdot,\boldsymbol{y})\rangle_{\mathcal{B}_\Phi^p(\mathbb{R}^d)} = (Eh)(\boldsymbol{y}) = h(\boldsymbol{y}),$$

for all $h \in \mathcal{B}_\Phi^p(\Omega)$ and all $\boldsymbol{y} \in \Omega$, where E is the extension operator from $\mathcal{B}_\Phi^p(\Omega)$ into $\mathcal{B}_\Phi^p(\mathbb{R}^d)$ such that $Eh|_\Omega = h$ and $\|Eh\|_{\mathcal{B}_\Phi^p(\mathbb{R}^d)} = \|h\|_{\mathcal{B}_\Phi^p(\Omega)}$. Since $K(\boldsymbol{x},\cdot)|_\Omega \in \mathcal{B}_\Phi^p(\Omega)$ for all $\boldsymbol{x} \in \mathbb{R}^d$, we can also obtain the other-side reproduction of $\mathcal{B}_\Phi^p(\Omega)$, i.e.,

$$\overline{\langle K(\boldsymbol{x},\cdot)|_\Omega, g\rangle}_{\mathcal{B}_\Phi^p(\Omega)} = \overline{\langle K(\boldsymbol{x},\cdot), g\rangle}_{\mathcal{B}_\Phi^p(\mathbb{R}^d)} = g(\boldsymbol{x}),$$

for all $g \in \mathcal{N}_0^\perp \equiv \mathcal{B}_\Phi^p(\Omega)'$. Therefore $\mathcal{B}_\Phi^p(\Omega)$ is a strong-sense RKBS with the reproducing kernel $K|_{\mathbb{R}^d\times\Omega}$.

Since $\mathcal{B}_\Phi^p(\mathbb{R}^d)$ is uniformly convex, [16, Theorem 5.2.24] provides that $\mathcal{B}_\Phi^p(\Omega) \equiv \mathcal{B}_\Phi^p(\mathbb{R}^d)/\mathcal{N}_0$ is uniformly convex. We also know that $\mathcal{B}_\Phi^p(\mathbb{R}^d)' \equiv \mathrm{L}_q(\mathbb{R}^d;\mu)$ is uniformly convex and $\mathcal{N}_0^\perp$ is a closed subspace of $\mathcal{B}_\Phi^q(\mathbb{R}^d) \equiv \mathcal{B}_\Phi^p(\mathbb{R}^d)'$ by [16, Proposition 1.10.15]. Combining with [16, Proposition 5.1.20 and 5.4.5], we can also check that $\mathcal{B}_\Phi^p(\Omega)$ is smooth. □

Remark 4.8. When $p = 2$, then we know that $\mathcal{B}_\Phi^2(\Omega)$ is a Hilbert space by Theorem 4.3. Thus its dual space and itself are isometrically isomorphic such that its reproducing kernel becomes $K|_{\Omega\times\Omega}$. Since $\mathcal{B}_\Phi^2(\mathbb{R}^d) = \mathcal{N}_0 \oplus \mathcal{N}_0^\perp$, we can determine that $\left\{g|_\Omega : g \in \mathcal{N}_0^\perp\right\} = \mathcal{B}_\Phi^2(\Omega)$ and $\|g\|_{\mathcal{B}_\Phi^2(\mathbb{R}^d)} = \|g|_\Omega\|_{\mathcal{B}_\Phi^2(\Omega)}$ for all $g \in \mathcal{N}_0^\perp$ which implies that $\mathcal{B}_\Phi^2(\Omega) \equiv \mathcal{N}_0^\perp \equiv \mathcal{B}_\Phi^2(\Omega)'$ and $\mathcal{B}_\Phi^2(\Omega)$ has the inner product

$$(h_1,h_2)_{\mathcal{B}_\Phi^2(\Omega)} = \langle h_1,h_2\rangle_{\mathcal{B}_\Phi^2(\Omega)} = \langle Eh_1,Eh_2\rangle_{\mathcal{B}_\Phi^2(\mathbb{R}^d)} = (Eh_1,Eh_2)_{\mathcal{B}_\Phi^2(\mathbb{R}^d)},$$

for all $h_1,h_2 \in \mathcal{B}_\Phi^2(\Omega)$. Therefore $\mathcal{B}_\Phi^2(\Omega)$ is a RKHS. Moreover, since $K(\cdot,\boldsymbol{y}) \in \mathcal{N}_0^\perp$ for any $\boldsymbol{y} \in \Omega$, we have $E\left(K(\cdot,\boldsymbol{y})|_\Omega\right) = K(\cdot,\boldsymbol{y})$. This shows that $K|_{\Omega\times\Omega}$ is a reproducing kernel of $\mathcal{B}_\Phi^2(\Omega)$. This conclusion is the same as in [3, Theorem 6].

If the RKBS is even a Hilbert space, then we can choose an equivalent function space of its dual as itself such that its reproducing kernel has symmetric domains. The difficulty to find an equivalent function function space of the dual of RKBS, which is defined on the same domain of the RKBS, causes the domains of its reproducing kernel to be nonsymmetric. Theorem 4.3 and 4.7 provide us the examples of symmetric and nonsymmetric reproducing kernels of RKBSs, respectively.

If the subset Ω is a *regular* domain, then the definition of weak derivatives (see [1, Section 1.62]) provides that $f|_\Omega \in W_p^m(\Omega)$ and $\|f|_\Omega\|_{W_p^m(\Omega)} \leq \|f\|_{W_p^m(\mathbb{R}^d)}$ for all $f \in W_p^m(\mathbb{R}^d)$, where $W_p^m(\Omega)$ is the L_p-based Sobolev space of order m. Now we use the imbeddings of $\mathcal{B}_\Phi^p(\mathbb{R}^d)$ to derive the imbeddings of $\mathcal{B}_\Phi^p(\Omega)$. We fix any $h \in \mathcal{B}_\Phi^p(\Omega)$. According to Corollary 4.5, we have

$$\|h\|_{W_p^m(\Omega)} \leq \|Eh\|_{W_p^m(\mathbb{R}^d)} \leq C\,\|Eh\|_{\mathcal{B}_\Phi^p(\mathbb{R}^d)} = C\,\|h\|_{\mathcal{B}_\Phi^p(\Omega)}, \quad h \in \mathcal{B}_\Phi^p(\Omega) \subseteq W_p^m(\Omega),$$

for some positive constant C independent on h.

Corollary 4.8. *Let $\mathcal{B}^p_\Phi(\mathbb{R}^d)$ with $p \geq 2$ and $m > pd/q - d/q$ be defined in Corollary 4.5. Here q is the conjugate exponent of p. Suppose that $\Omega \subseteq \mathbb{R}^d$ is regular. Then $\mathcal{B}^p_\Phi(\Omega)$ defined in Theorem 4.7 is imbedded into the L_p-based Sobolev space of order m, $W^m_p(\Omega)$, i.e.,*

$$\|h\|_{W^m_p(\Omega)} \leq C\, \|h\|_{\mathcal{B}^p_\Phi(\Omega)}, \quad h \in \mathcal{B}^p_\Phi(\Omega) \subseteq W^m_p(\Omega),$$

for some positive constant C independent on h.

5 Examples for Matérn Functions

[9, Example 5.7] and [22, Example 4.4] show that *Matérn functions* (Sobolev splines) with shape parameter $\theta > 0$ and degree $n > d/2$

$$G_{\theta,n}(\boldsymbol{x}) := \frac{2^{1-n-d/2}}{\pi^{d/2}\Gamma(n)\theta^{2n-d}}(\theta\,\|\boldsymbol{x}\|_2)^{n-d/2}K_{d/2-n}(\theta\,\|\boldsymbol{x}\|_2), \quad \boldsymbol{x} \in \mathbb{R}^d,$$

is a positive definite function on $\mathbb{R}^d$, where $t \mapsto K_\nu(t)$ is the modified Bessel function of the second kind of order ν and $t \mapsto \Gamma(t)$ is the Gamma function. Moreover, $G_{\theta,n}$ is a full-space Green function of differential operator $L_{\theta,n} := \left(\theta^2 I - \Delta\right)^n$, i.e., $L_{\theta,n}G_{\theta,n} = \delta_{\boldsymbol{0}}$. The Fourier transform of $G_{\theta,m}$ has the form

$$\hat{G}_{\theta,n}(\boldsymbol{x}) = \left(\theta^2 + \|\boldsymbol{x}\|_2^2\right)^{-n}, \quad \boldsymbol{x} \in \mathbb{R}^d.$$

Let $1 < q \leq 2 \leq p < \infty$ with $p^{-1} + q^{-1} = 1$ such that $nq/p > d/2$ and $m := 2n/q$. Since $\hat{G}^{q/p}_{\theta,n} \in \mathrm{L}_1(\mathbb{R}^d)$, Theorem 4.3 provides that $\mathcal{B}^p_{G_{\theta,n}}(\mathbb{R}^d)$ is a RKBS on $\mathbb{R}^d$ with a reproducing kernel $K_{\theta,n}(\boldsymbol{x},\boldsymbol{y}) = G_{\theta,n}(\boldsymbol{x}-\boldsymbol{y})$. We can also check that there are two positive constants C_1, C_2 such that

$$C_1\left(1 + \|\boldsymbol{x}\|_2^2\right)^{-m/2} \leq \hat{G}_{\theta,n}(\boldsymbol{x})^{1/q} \leq C_2\left(1 + \|\boldsymbol{x}\|_2^2\right)^{-m/2}. \quad \boldsymbol{x} \in \mathbb{R}^d,$$

According to Corollary 4.5 and 4.8, the RKBS $\mathcal{B}^p_{G_{\theta,n}}(\mathbb{R}^d)$ is imbedded into $W^m_p(\mathbb{R}^d)$ and the RKBS $\mathcal{B}^p_{G_{\theta,n}}(\Omega)$ is imbedded into $W^m_p(\Omega)$ for any regular domain Ω of $\mathbb{R}^d$.

In particular, we let $p := 4$. Then $\hat{G}^3_{\theta,n} = \hat{G}_{\theta,3n}$. According to the discussion of Corollary 4.6 and Remark 4.7, the optimal solution of the support vector machine

$$\min_{f \in \mathcal{B}^4_{G_{\theta,n}}(\mathbb{R}^d)} \sum_{j=1}^N L(\boldsymbol{x}_j, y_j, f(\boldsymbol{x}_j)) + \Sigma\left(\|f\|_{\mathcal{B}^4_{G_{\theta,n}}(\mathbb{R}^d)}\right),$$

has the empirical representation

$$s_{D,L,\Sigma}(\boldsymbol{x}) = \sum_{j,k,l=1}^{N,N,N} c_j\overline{c_k}c_l G_{\theta,3n}\left(\boldsymbol{x} - \boldsymbol{x}_j + \boldsymbol{x}_k - \boldsymbol{x}_l\right) = \sum_{j,k,l=1}^{N,N,N} c_j\overline{c_k}c_l K_{\theta,3n}\left(\boldsymbol{x} + \boldsymbol{x}_k, \boldsymbol{x}_j + \boldsymbol{x}_l\right), \quad \boldsymbol{x} \in \mathbb{R}^d,$$

and its coefficients $\boldsymbol{c} = (c_1, \cdots, c_N)^T$ are solved by the following minimal problem

$$\min_{\boldsymbol{c} \in \mathbb{C}^N} \sum_{i=1}^N L\left(\boldsymbol{x}_i, y_i, \sum_{j,k,l=1}^{N,N,N} c_j\overline{c_k}c_l K_{\theta,3n}\left(\boldsymbol{x}_i + \boldsymbol{x}_k, \boldsymbol{x}_j + \boldsymbol{x}_l\right)\right) + \Sigma\left(\sum_{i,j,k,l=1}^{N,N,N,N} \overline{c_i}c_j\overline{c_k}c_l K_{\theta,3n}\left(\boldsymbol{x}_i + \boldsymbol{x}_j, \boldsymbol{x}_k + \boldsymbol{x}_l\right)\right)^{1/2},$$

where the loss function L and Σ are defined as in Corollary 4.6. More general, when $p = 2l$ is even, then the support vector solution $s_{D,L,\Sigma}$ in $\mathcal{B}^{2l}_{G_{\theta,n}}(\mathbb{R}^d)$ is a linear combination of the product groups of the reproducing kernel bases, i.e.,

$$s_{D,L,\Sigma}(\boldsymbol{x}) = \sum_{\boldsymbol{k}\in\mathcal{G}^N_{p-1}} \prod_{j=1}^{l} c_{k_{2j-1}} \prod_{j=1}^{l-1} \overline{c_{k_{2j}}} K_{\theta,(p-1)n}\left(\boldsymbol{x} + \sum_{j=1}^{l-1} \boldsymbol{x}_{k_{2j}}, \sum_{j=1}^{l} \boldsymbol{x}_{k_{2j-1}}\right), \quad \boldsymbol{x} \in \mathbb{R}^d,$$

where $\boldsymbol{k} := (k_1, \cdots, k_{p-1})^T$ and $\mathcal{G}^N_{p-1} := \left\{\boldsymbol{k} \in \mathbb{N}^{p-1} : 1 \le k_j \le N \text{ for all } j = 1, \ldots, p-1\right\}$.

The Matérn functions has been applied into the fields of statistical learning (see [14]). The new discover of Matérn functions could help the people to create a new numerical tool for support vector machines in RKBS.

6 General Support Vector Machines

In this paper, we can review the empirical support vector machine (4.2) by the empirical measure $\mathbb{P}_D := \sum_{k=1}^N \delta_{(\boldsymbol{x}_k,y_k)}$ of the observations data $D := \{(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_N, y_N)\}$, i.e.,

$$\min_{f\in\mathcal{B}} \int_{\Omega_1\times\mathbb{C}} L(\boldsymbol{x}, y, f(\boldsymbol{x}))\mathrm{d}\mathbb{P}_D(\boldsymbol{x}, y) + \Sigma\left(\|f\|_{\mathcal{B}}\right).$$

We can generalize the empirical measure to the continuous probability measure $\mathbb{P}_C$ defined on $\Omega_1 \times \mathbb{C}$, e.g., the classification probability $\mathbb{P}_C(\boldsymbol{x}, y) := \eta(y|\boldsymbol{x})\mu(\boldsymbol{x})$, where μ is a continuous probability measure on Ω_1 and $\eta(\cdot|\boldsymbol{x})$ is a classification conditional probability measure on $\mathbb{C}$ given by

$$\eta(y|\boldsymbol{x}) := \begin{cases} \kappa(\boldsymbol{x}), & y = 1, \\ 1 - \kappa(\boldsymbol{x}), & y = -1, \\ 0, & \text{otherwise}, \end{cases}$$

with $\kappa(\boldsymbol{x}) \in [0, 1]$ for all $\boldsymbol{x} \in \Omega_1$. We can also use this probability measure $\mathbb{P}_C$ to set up a general support vector machine, i.e.,

$$\min_{f\in\mathcal{B}} \int_{\Omega_1\times\mathbb{C}} L(\boldsymbol{x}, y, f(\boldsymbol{x}))\mathrm{d}\mathbb{P}_C(\boldsymbol{x}, y) + \Sigma\left(\|f\|_{\mathcal{B}}\right).$$

In our future research work, we will try to obtain the representation of general support vector solutions similar as in [19, Theorem 5.8]. Its numerical experiments will appear in our later research papers.

References

[1] R. A. Adams and J. J. F. Fournier, *Sobolev Spaces*. vol. 140 of pure and applied mathematics, Elservier/Academic Press, Amsterdam, 2003.

[2] E. Alpaydin, *Introduction to Machine Learning*. MIT Press, 2010.

[3] A. Berlinet and C. Thomas-Agnan, *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Kluwer Academic Publishers, 2004.

[4] M. D. Buhmann, *Radial Basis Functions: Theory and Implementations*. vol. 12 of Cambridge monographs on applied and computational mathematics, Cambridge University Press, 2003.

[5] J. F. Erickson, *Generalized native spaces*. Ph.D. thesis, Illinois Institute of Technology, 2007.

[6] J. F. Erickson and G. E. Fasshauer, *Generalized native spaces*. Approximation Theory XII: San Antonio 2007, Mod. Methods Math., 2008, pp. 133–142.

[7] G. E. Fasshauer, *Meshfree Approximation Methods with* MATLAB. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2007.

[8] G. E. Fasshauer, *Positive definite kenrels: past, present and future*. Dolomite Research Notes on Approximation, vol. 4, 2011, pp. 21–63.

[9] G. E. Fasshauer and Q. Ye, *Reproducing kernels of generalized Sobolev spaces via a green function approach with distributional operators*. Numer. Math., vol. 119, 2011, pp. 585–611.

[10] G. E. Fasshauer and Q. Ye, *Reproducing kernels of Sobolev spaces via a Green kernel approach with differential operators and boundary operators*. Adv. Comput. Math., 2011, DOI: 10.1007/s10444-011-9264-6.

[11] J. R. Giles, *Classes of semi-inner-product spaces*. Trans. Amer. Math. Soc., vol. 129, No. 3, 1967, pp. 436–446.

[12] L. Hörmander, *The Analysis of Linear Partial Differential Operators I*. Classics in Mathematics, Springer-Verlag, 2003.

[13] R. C. James, *Orthogonality and linear functionals in normed linear spaces*. Trans. Amer. Math. Soc., vol. 61, 1947, pp. 265–292.

[14] M. Jun and M. L. Stein, *Nonstationary covariance models for global data*. Ann. Appl. Stat., vol. 2, no. 4, 2008, pp. 1271–1289.

[15] G. Lumer, *Semi-inner-product spaces*. Trans. Amer. Math. Soc., vol. 100, 1961, pp. 29–43.

[16] R. E. Megginson, *An Introduction to Banach Space Theory*. graudate texts in mathematics, Springer-Verlag, 1998.

[17] G. Song and H. Zhang, *Reproducing kernel Banach spaces with the l^1 norm II: Error analysis for regularized least square regression*. Neural Comput., vol. 23, no. 10, 2011, pp. 2713–2729.

[18] G. Song, H. Zhang and F. J. Hickernell, *Reproducing kernel Banach spaces with the l^1 norm*. Appl. Comput. Harmon. Anal., to appear, 2012.

[19] I. Steinwart and A. Christmann, *Support Vector Machines*. information science and statistics, Springer, New York, 2008.

[20] G. Wahba, *Spline Models for Observational Data*. vol. 59 of CBMS-NSF regional conference series in applied mathematics, SIAM, 1990.

[21] H. Wendland, *Scattered Data Approximation*. vol. 17 of Cambridge monographs on applied and computational mathematics, Cambridge University Press, 2005.

[22] Q. Ye, *Analyzing reproducing kernel approximation methods via a Green function approach*. Ph.D. thesis, Illinois Institute of Technology, 2012.

[23] H. Zhang, Y. Xu and J. Zhang, *Reproducing kernel Banach spaces for machine learning*. J. Mach. Learn. Res., vol. 10, 2009, pp. 2741–2775.

Gregory E. Fasshauer, Fred J. Hickernell
Department of Applied Mathematics, Illinois Institute of Technology, Chicago, Illinois 60616
E-mail address: fasshauer@iit.edu, hickernell@iit.edu

Qi Ye
Mathematics Department, Syracuse University, Syracuse, NY 13244
E-mail address: qiye@syr.edu

Yuesheng Xu
Mathematics Department, Syracuse University, Syracuse, NY 13244
E-mail address: yxu06@syr.edu